**PS-1: LLMs to Detect Vulnerability in Open-Source Software**

**Problem Statement:**

**Development of a LLM based tool for Vulnerability (Malicious Code) Detection in Source Code of Software (especially open source software) and generating mitigating measures**

**PS Description:**

“Development of a LLM based tool for Vulnerability (Malicious Code) Detection in Source Code of Software (especially open source software) and generating mitigating measures”. This includes extraction of Vulnerabilities, dependencies, list the CVEs, CWEs in vulnerable code in open source data bases and source code of vulnerable software etc. and suggest automated mitigation measures.

**PS-2: Phishing Detection**

**Problem Statement:**

**AIML based solution capable of detecting phishing domains and pages for specific target organisations.**

**PS Description:**

Phishing attacks remain one of the most persistent and damaging threats in the cybersecurity landscape. Cybercriminals often register domains that closely resemble those of well-known organizations in order to deceive users into revealing sensitive information such as login credentials, personal identification data, financial data etc.

Through the AI Grand Challenge, it is proposed to develop a platform to monitor, identify and alert phishing/suspected domains/urls targeting nation’s Critical Sector Entities (CSEs). The platform to continuously and automatically scan various domains/urls such as newly created TLD domains, domains on Hosted Infra e.g. ngrok etc, Social Media platforms, Open Source etc. and may use a combination of techniques such as domain similarity analysis, web content inspection, DNS/WHOIS metadata evaluation etc. to classify domains/urls as Phishing (having CSE look alike contents) or Suspected (parked domains/urls without lookalike content). In cases where domains/urls do not initially host phishing content but the domain/url exhibits high similarity to legitimate domain, they will be flagged as Suspected and entered into a long-term, configurable monitoring / alerting workflow. All phishing domain/url alert/report to include domain metadata, screenshots, phishing indicators, external verification data, related CSE name etc.

By integrating detection, classification, continuous monitoring and automated reporting into a single framework, the platform to provide an end- to-end modular and scalable solution for defending against evolving phishing threats targeting CSEs.

**PS-3: Visual Search, Retrieval and Detection in Satellite Imagery**

**Problem Statement:**

**To develop a system to automatically search, identify and detect objects in a wide range of satellite images, producing labelled datasets that can be further trained and fine-tuned to object detection models directly.**

**PS Description:**

The spate of the recent years has witnessed an explosion in satellite imagery that has opened tremendous opportunities for variety of applications like urban planning, environmental monitoring, disaster response etc. Effective use of such data, though, depends extensively on large-scale and accurate annotated datasets, especially for training object detection models to automatically detect and localize certain objects like buildings, vehicles etc. Manual annotation of satellite images is a very time- consuming, labor-intensive and expensive task because of the varying resolution of images, the complexity of the scenes and the requirement of expert knowledge to differentiate between visually similar objects. Hence, there is a need to create a scalable, efficient and automatic solution that can search, retrieve and detect any given object/target features in multispectral satellite imagery datasets. The underlying goal of this problem is to develop a system or framework that accepts one or few image sample/s of any object/feature class and can automatically search, identify and detect that object in a wide range of satellite images, producing labelled datasets that can be further trained and fine-tuned to object detection models directly. The solution will allow users to provide input by uploading an image chip of a feature/object of interest or drawing a box around a feature/object of interest over a selected imagery. The AI Accelerator of RAIC Labs (<https://raiclabs.com/products/ai-accelerator/>), orcahunter modules of BlackShark.ai (<https://blackshark.ai/orcahuntr/>) can be used as references.

**PS-4: Retrieval Augmented Generation based Question and Answering System**

**Problem Statement:**

**Retrieval Augmented Generation based Question and Answering System**

**PS Description:**

The PS entails development of an information correlation across multiple documents to generate summaries that are grounded in evidence. The RAG solution should combine semantic retrieval and generative reasoning.

Objectives:

1) Retrieval system for relevant documents retrieval from a diverse documents corpus.

2) Generate fluent, factually grounded and explainable natural language responses.

3) Minimize hallucination & ensure all claims are traceable to source content

**PS-5: Intelligent Multi-lingual Document Understanding**

**Problem Statement:**

**Digitization/ Text Extraction for multi-lingual, noisy documents while preserving document Layout including hand written documents.**

**PS Description:**

In today’s global and digitized world, documents are far more than just sequences of text. They are structured, visually rich, and multilingual—ranging from legal contracts and academic papers to business reports, government forms, and presentation decks. These documents exist in diverse formats such as Word documents (DOCX), PDFs, PowerPoint slides (PPT), and scanned images including handwritten documents, often containing mixed scripts (e.g., English-Arabic or Hindi-English). Modern AI systems need to read, understand, and generate structured outputs from these documents.

The Challenge

* Multilingual layout-aware document parsing across scripts, formats, and writing directions.
* Accurately extracting structured information from documents while preserving:
* Visual hierarchy (headings, sections),
* Semantic grouping (form fields, captions, references),
* Layout fidelity (table structures, image alignment, reading order),
* Embedded elements like charts, plots, maps, and figures.
* Representing the extracted content in a standardized, machine-friendly yet human-readable format. A document may contain the following:
* Plain text (headers, paragraphs)
* Table
* Image
* Map
* Charts

The solution should be able to localise, classify and convert the above components into **natural language text and provided the output in json and markdown format with languages identified**.

**PS-6: Language Agnostic Speaker Identification / Verification & Diarisation; and subsequent Transcription & Translation System**

**Problem Statement:**

**Development of a robust System for Language-Agnostic Speaker Identification/Verification and Diarization; and subsequent Transcription & Translation of the multilingual and code-switched audio.**

**PS Description:**

We aim to develop an integrated audio processing system that transforms spoken audio into structured, multilingual and textual insights. Given an input audio file, the system will perform the following tasks :-

**Speaker Identification:**

Match each speaker segment to a known identity when enrolment data is available.

**Speaker Diarization:**

Segment the audio by identifying boundary between different speakers.

**Language Identification:**

Detect the language spoken in each segment, supporting multilingual and code-switched audio.

**Automatic Speech Recognition:**

Convert each speaker’s speech into accurate text in spoken script.

**Neural Machine Translation:**

Translate the transcribed text into English language, preserving speaker level segmentation.

**The solution should be able to handle:**

1. Sample rate – 8k – 48 k
2. Bit depth : 4 - 32 bits
3. File types : wav, mp3, ogg, flac etc.
4. SNR : -5 to 20 db

**PS-7: Password Extraction & Decryption**

**Problem Statement:**

**Extraction and Decryption of Passwords from password protected documents, disks and compressed files for retrieval of Forensic evidence.**

**PS Description:**

Traditional password analysis and cracking methods, such as brute force and dictionary attacks, are becoming less effective, time-consuming against complex and well-secured passwords. Further with the proliferation of data breaches and the advancement of computing power, traditional password cracking methods have become less effective. Advancements in AI/ML techniques offer opportunities to enhance password analysis and cracking processes by leveraging pattern recognition, probabilistic modelling, and predictive analytics.

This critical challenge can be addressed by harnessing the power of Artificial Intelligence (AI) and Machine Learning (ML) techniques for password analysis and cracking. The goal of this problem is to Develop AI/ML models to analyse password datasets, extract patterns, and identify common characteristics. Utilize these models to generate more effective password cracking strategies, and enhance password cracking efficiency, accuracy and maximise the success rates using machine learning algorithms. **This AI/ML-driven solution will go beyond conventional techniques, providing a new paradigm for password analysis and cracking.**

**PS-8: Emitter Location Enhancement**

**Problem Statement:**

**Develop a Solution to provide cross-sensor localization enhancement based on reduction in the Ellipse Error Probable.**

**PS Description:**

The Problem Statement pertains to finding precise geographic location of a signal source. This is achieved by the following:

* Collecting RF(Radio Frequency) signals from various geographic positions
* Estimating direction of the received RF signal along with other parameters
* Applying Triangulation method, to check convergence of received signal direction to a common point.

Normally, triangulation of collected RF signal gives a rough idea of signal source location and an area is defined where the probability of locating the signal source is highest. This area could be large/ small which depends upon the accuracy of signal collected by a receiving system.

The accuracy of the receiving system requires constant calibration of the hardware, stability of platform where the system is installed and other RF interference. With these unavoidable circumstances, the location area calculated by the system are larger and accuracy of signal location suffers. Improvement can be made in terms of post processing of data by applying different estimation algorithm/ model which could enhance the accuracy of signal location.

**PS-9: Maritime Domain Awareness**

**Problem Statement:**

**Solution for Detection & Classification of Objects of Interest in wide swath SAR/EO imagery and multi-sensor data correlation.**

**PS Description:**

An automated tool to undertake the following:

* Vessel detection on medium resolution of satellite imagery in open sea
* Correct estimation of Length, Width and Heading of the detected ships
* Classification of these detections into commercial vessels, vessels of interest based and other classes based on a training dataset developed in consultation with user.
* Correlation of these detections with the following data sets: -
* Automatic Identification System (AIS) data
* RF Geolocation data
* Other time-series data sets provided in the form of csv or postgres data base
* Vessel path interpolation/prediction in open sea based on past data

**PS-10: Change Detection in Satellite Imagery**

**Problem Statement:**

**Development of AI Solutions for Automated Detection and Analysis of man- made Changes of large landmass in Satellite Imagery over time.**

**PS Description:**

Availability of every day satellite imageries giving carpet coverage of entire landmass of regions of our interest has given the capability to continuously monitor any new developments. But it is impossible to handle this much volume of satellite image using trained human resource to analyze because of the scale and volume. Hence automation of change detection on such datasets using AI/ML can be of great operational usage to detect any new man-made developments. AI/ML based change detection models can reduce false alarms in such change detection.

Availability of large volume of satellite imageries of medium spatial resolution and high temporal resolution, provides enhanced earth observation capabilities. To handle this voluminous data, automated techniques are required to detect and monitor changes on a global scale. Therefore, an automated change detection solution using AI/ML is expected to be developed, to detect man-made changes.

**PS-11: Hyperspectral Anomaly Detection**

**Problem Statement:**

**AI/ML Based Spectral-Spatial Anomaly Detection in Hyperspectral and Thermal (IR) Remote Sensing Datasets**

**PS Description:**

Anomaly detection in hyperspectral and thermal remote sensing datasets is a critical task in remote sensing applications such as environmental monitoring, mining, precision agriculture, and disaster management. These datasets offer rich spectral and/or thermal information, allowing the identification of materials, objects, or behaviours that deviate from the norm.

Hyperspectral data cubes contain images in hundreds of contiguous spectral bands, where each pixel has a spectral signature, making it possible to distinguish materials. Deep learning based spectral-spatial anomaly detection results (suppressing natural anomalies) would eventually enhance the data interpretability and material identification.

Similarly, slight temperature changes due to human activities such as industrial operations, fire incidents, machine operations etc. in thermal datasets could be detected using deep learning based thermal anomaly detection and subsequent material identification referencing material thermal emissivity.

Objectives:

1) Deep learning model based manmade anomaly detection using satellite hyperspectral images (400-2500 nm).

2) Characterization of materials in detected anomalies (pixels) using spectral signature library matching techniques.

3) Thermal anomalies detection associated with human activities using deep learning techniques.

4) Characterization and identification of thermal anomalies using thermal emissivity of various materials.

**PS-12: Underwater Domain Awareness**

**Problem Statement:**

**AI Solution for Detection & Classification of objects in Underwater Acoustic Data.**

**PS Description:**

To detect anomalies (in the background of ambient noise) in an audio file (.wav format), extract the duration of anomalies (presence of objects), extract features of the object and classify the object into different classes.